



Audio Engineering Society Convention Paper 10069

Presented at the 145th Convention
2018 October 17 – 20, New York, NY, USA

This convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A Statistical Metric for Stability in Instrumental Vibrato

Sarah R. Smith¹ and Mark F. Bocko¹

¹University of Rochester, Department of Electrical and Computer Engineering

Correspondence should be addressed to Sarah Smith (sarahsmith@rochester.edu)

ABSTRACT

When instrumentalists perform with vibrato, they add a quasi-periodic frequency modulation to the note. Although this modulation is rarely purely sinusoidal, many methods for vibrato parameterization focus exclusively on the rate and depth of the frequency modulation, with less attention given to measuring how a performer's vibrato changes over the course of a note. In this paper, we interpret the vibrato trajectories as instantiations of a random process which can be characterized by an associated autocorrelation function and power spectral density. From these distributions, a coherence time can be estimated that describes the stability of the vibrato within a note. This metric can be used to characterize individual performers as well as for resynthesizing vibratos of different styles.

1 Introduction

Vibrato is a commonly used musical technique where the performer introduces a frequency modulation to the note for expressive effect. While a frequency modulation is always present, and dominates bowed string vibrato, large modulations of amplitude and spectra are also observed in many cases [1, 2, 3]. Furthermore, a trained performer will adjust both the rate and width of their vibrato depending on the musical context. As such, most analysis of vibrato focuses on these parameters when characterizing different instruments or players [4, 2].

In this context, the vibrato is often assumed to be sinusoidal, or nearly so. However, natural vibratos are never perfectly periodic and generally contain some random variations in rate, width, and shape. These variations can appear both during a note, as the

vibrato waveform evolves and the note is shaped, or between different notes on the instrument or phrase. An example of these differences is given in Figure 1. Here, three different vibrato waveforms are shown, all of which have similar rate and width, but differ in their evolution. In particular, the waveform of the middle example is seen to change about half way through the note, becoming smaller and more rounded.

In order to characterize these changes, we develop a method for describing and comparing the vibratos of different players that is independent of the rate and width of the modulation. This parameter, which we call stability, corresponds to the length of time over which a given vibrato sample remains consistent. In particular, this paper extends and supplements existing methods of vibrato analysis to include an explicit treatment of vibrato as a stochastic as opposed to deterministic process. In addition to presenting and defining

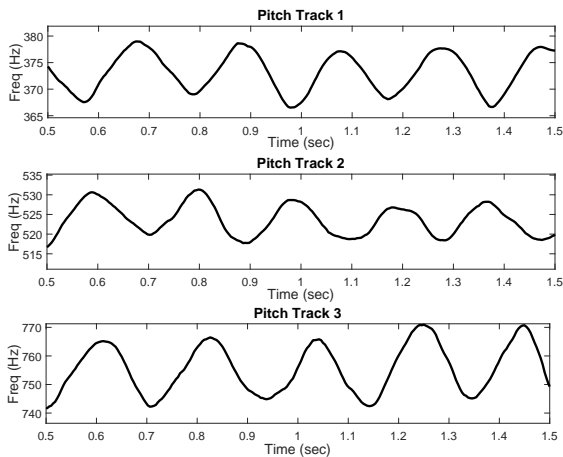


Fig. 1: Pitch tracks from three samples of vibrato are shown. All three vibratos change over the course of the note, but in slightly different ways.

a parameterization scheme and quantitative metric for vibrato stability, it demonstrates how this metric can be used to compare vibrato from different performers across a varied database of musical tones.

2 Audio Feature Extraction

The first stage of vibrato analysis is to extract the vibrato parameter tracks from the recorded audio. Before analysis, the steady state portion of the tone is isolated by removing the first and last 15 percent of the note. Depending on the context, a number of different parameter tracks may be used to characterize the vibrato. For instance, a pitch track can be estimated using the Yin algorithm [5]. In addition, the spectral centroid and overall amplitude tracks can be useful for characterizing many instruments, particularly woodwinds [6, 3]. Although the majority of this paper focuses on the analysis of pitch trajectories, a similar analysis could be performed on other trajectories.

2.1 Isolation of Vibrato

The raw parameter tracks often include variations from sources other than vibrato. In particular, the pitch tracks often contain minor intonation adjustments and the amplitude tracks can be dominated by an overall loudness envelope. In order to analyze the vibrato independently of other parameter variations, it is necessary to separate

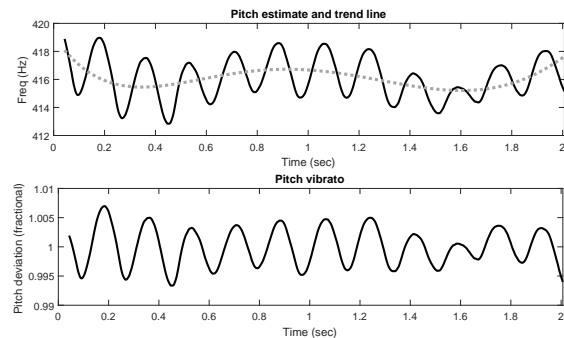


Fig. 2: The original pitch estimate (black, solid) and polynomial trend line (grey, dotted) for a vibrato sample are shown at top. The vibrato track (shown at bottom) is the difference between the raw track and the trend.

the faster vibrato variations from the underlying trends. Mathematically, this amounts to a decomposition of the raw parameter track $p_r[t]$ into a vibrato component $p_v[t]$ and a trend component $p_t[t]$ as given in Equation 1.

$$p_r[t] = p_v[t] + p_t[t] \quad (1)$$

In the following analysis, the trend $p_t[t]$ is estimated by fitting a 5th order polynomial to the raw parameter track, however the polynomial order could be adjusted depending on the consistency of the performer and the amount of non-vibrato deviations that are present. While this trend could also be estimated by low pass filtering the raw parameter track, the filters needed to isolate a sufficiently narrow passband of only a few Hz generally have correspondingly long filter lengths in the time domain, and therefore require a method to explicitly handle the ends of the parameter tracks. In addition to avoiding explicit endpoint corrections, a polynomial fit also provides a compact parametric representation of the trend which can be easily manipulated or removed during resynthesis.

An example of this process is shown in Figure 2. Although a prominent vibrato can be seen in the upper (raw) pitch track, there is also a slowly varying trend indicated with a dotted line. In the lower figure, the underlying trend has been removed to generate the vibrato waveform.

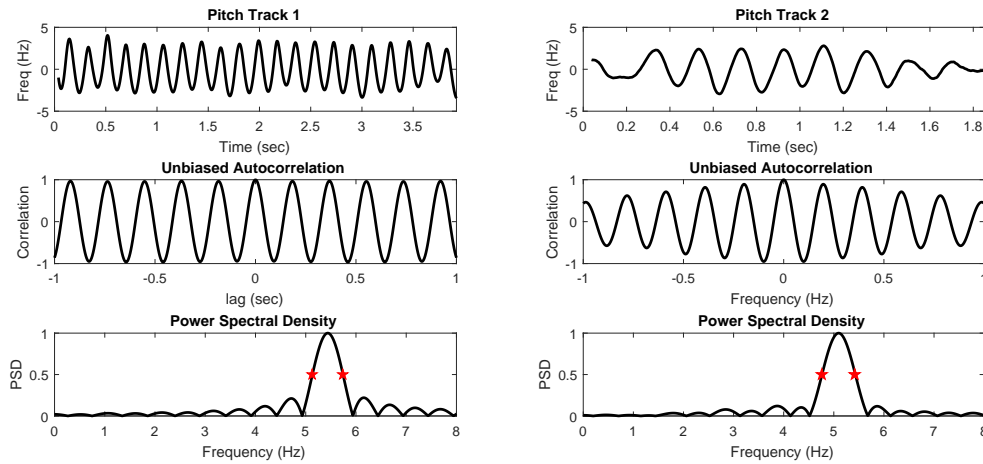


Fig. 3: Two vibrato tracks are shown (top) along with their corresponding autocorrelations (middle) and power spectral densities (bottom). The example at left consists of a very stable vibrato with a wide autocorrelation and narrow PSD. In contrast the right hand example shows a less stable vibrato characterized by a narrower autocorrelation and wide PSD.

3 Statistical Analysis of Vibrato

Once the vibrato tracks have been extracted and isolated from the audio recording, the statistical properties of the process can be estimated. Like many natural processes, a musician's vibrato contains some amount of unpredictability, and therefore cannot be described as a purely deterministic process.

Although many methods of analyzing vibrato identify global parameters like width and rate, less work has been done to characterize the random component of the modulation. By interpreting the vibrato parameter track as a single instantiation of a random process, it is possible to identify factors that are consistent between notes of the same player or differ significantly between individual performers or instruments. In this framework, the vibrato parameters are not estimated directly from the parameter tracks, but calculated from the corresponding autocorrelation and power spectral density functions.

3.1 Vibrato Autocorrelation

To characterize the time evolution of the vibrato, the autocorrelation function is used. The autocorrelation function $r_p[\tau]$ compares the original signal with itself

at different a series of offsets, τ as given in Equation 2. Therefore, a perfectly sinusoidal vibrato of infinite duration will have a sinusoidal autocorrelation with infinite width and a rate of oscillation equal to the vibrato rate.

$$r_p[\tau] = \sum_t p_v[t]p_v[t + \tau] \quad (2)$$

In practice, however, the width of the autocorrelation is limited both by the finite duration of the tone and steadiness of the player. In order to compare notes of different lengths, we use the unbiased sample autocorrelation, $r_{ub}[\tau]$, as given in Equation 3, where T_{max} is the length of the vibrato trajectory. The unbiased sample autocorrelation provides an estimate of the autocorrelation function of the underlying process, which can then be used to characterize the vibrato and compare multiple notes within a set.

$$r_{ub}[\tau] = \frac{1}{T_{max} - |\tau|} r_p[\tau] = \frac{1}{T_{max} - |\tau|} \sum_{t=0}^{t=T_{max}} p_v[t]p_v[t + \tau] \quad (3)$$

This form removes the implicit Bartlett window that occurs when the raw autocorrelation is calculated for

a signal of finite duration [7]. However, this unbiasing does not fully compensate for the differing lengths, since it effectively replaces the implied Bartlett window with a rectangular window of the same length. As such, the autocorrelations must all be truncated to the same length when comparing tones of differing lengths in the frequency domain. While this cannot fully eliminate the effects of the window, it ensures that all samples in the database are treated uniformly and that the resulting statistical comparisons will not be compromised by differing signal lengths.

3.2 Vibrato Power Spectral Density

The power spectral density (PSD) is defined as the discrete time Fourier transform of the unbiased autocorrelation function, and can be calculated from the unbiased sample autocorrelation as defined in Equation 4.

$$S(\omega) = \sum_{\tau} r_{ub}[\tau] e^{-j\omega\tau} \quad (4)$$

Figure 3 shows an example of these calculations for two different vibrato samples. In the first (at left) example, the vibrato remains consistent throughout a long note, and therefore has a wide autocorrelation that does not decay much over the range of lags shown. In contrast, the second example (at right) is less stable, and the autocorrelation envelope decays noticeably. These differences can also be seen in the frequency domain, where the wider autocorrelation of the first example produces a correspondingly narrower power spectral density.

4 Vibrato Parameters

A range of vibrato parameters can be calculated from the functions defined in the previous section. These include statistical definitions of the vibrato width and rate as well as metrics for the stability of vibrato within a single note and consistency of a player across a number of tones.

4.1 Vibrato Width

The vibrato width is calculated directly from the standard deviation of the vibrato parameter track. For a pitch track, the standard deviation of the vibrato in Hz (σ_{Hz}) is converted to musical cents (1 semitone = 100 cents) using the relation given in Equation 5. This allows vibrato on different notes to be compared

and corresponds with both musical convention and the perception of pitch on a logarithmic scale.

$$Width = 1200 * \log_2\left(\frac{f_0 + \sigma_{Hz}}{f_0 - \sigma_{Hz}}\right) \quad (5)$$

A similar normalization is often desirable when analyzing other parameters. For example, when an amplitude trajectory is used, the raw amplitude values can be divided by the average amplitude to compare notes of different dynamics. The results can be presented either as fractional change, or converted to decibels as desired. Similarly for the spectral centroid, the raw centroid values in Hz may be divided by the fundamental pitch, producing a normalized value that is independent of the fundamental frequency.

4.2 Vibrato Rate

For a pitch track, the vibrato rate is taken as the location of the peak of the power spectral density function. However, for other parameter tracks, the peak finding algorithm may need to be restricted to a smaller range, since these trajectories often contain large variations at integer multiples of the vibrato rate, due to the acoustic resonances of the instrument [2].

4.3 Vibrato Stability

The features defined above, and considered elsewhere, are global parameters of the vibrato for a given note. When the vibrato changes over the duration of the note, they often represent the average parameters of the modulation. In contrast, we define stability as a measure of how consistent the vibrato is from one period to the next over the course of a note.

Mathematically, the stability is defined as the reciprocal of the vibrato bandwidth as defined by the full width at half max of the power spectral density, and given in Equation 6. Here, the factor of 2π in the numerator converts the bandwidth from radians per second to Hz when taking the inverse.

$$Stability = \frac{2\pi}{FWHM\{S(\omega)\}} \quad (6)$$

Conceptually, the stability estimates the length of time over which the vibrato process remains constant, and

closely parallels the concept of coherence time used in optics and other fields [8]. A larger stability value indicates a steadier, more regular vibrato, while a smaller value could correspond to intentional changes in the vibrato over the course of the note (such as changing the width or rate), or other inconsistencies in the vibrato waveform.

5 Ensemble Analysis of Vibrato Databases

In order to assess the utility of this metric, stability statistics were calculated for a database of vibrato tones played on many different instruments. The database used consisted of samples from the McGill University Master Samples, as well as the University of Iowa electronic music samples and real world computing database [9, 10, 11]. Although this analysis was run on the full database of string, woodwind, and brass instruments, this discussion highlights trends within the flute family instruments.

All three databases included multiple sets of flute tones with vibrato. The Iowa samples include treble, alto and bass flutes, each recorded by the same performer at different dynamic levels. The McGill samples include tones from one instrument of each type (piccolo, flute, alto flute, bass flute), although the performer is not specified. Finally, the real world computing database includes vibrato tones on two different treble flutes.

The stability values for this set of flute data is shown in Figure 4, with error bars denoting the standard deviation for each set of tones. From this data, some interesting patterns begin to emerge. For example, the Iowa database (blue, open circles) includes separate data at different dynamic levels. As such the flute family data includes flute, alto flute, and bass flute each recorded at ff and mf dynamics. These pairs of data sets have very similar vibrato statistics, which is unsurprising since the dynamic level is unlikely to heavily influence the performer's vibrato. Additionally, the database metadata tells us that all three of these instruments were recorded by the same performer, suggesting that this metric may be useful for characterizing the performance independent of the specific instrument.

This data, however, is only a small part of the analysis that can be done on these databases and is intended to

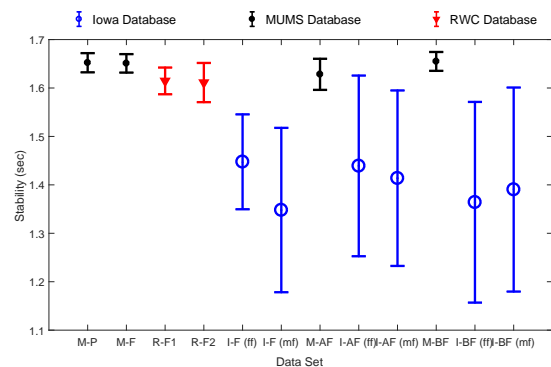


Fig. 4: Stability data is shown for a flute family vibrato across three databases. Data sets that are known to be performed by the same player on different instruments (blue open circles) have similar stability values.

motivate the relevance and applicability of the stability metric. When combined with other existing metrics for width and rate, the stability can provide a useful additional insight into performance patterns in vibrato.

6 Conclusions and Future Work

This paper presents a statistical framework for analyzing musical vibrato and introduces a metric to characterize the stability of vibrato over the course of a note. Some initial data is presented for a subset of tones in a large database which shows promise for future use of this metric. In addition to extending the current analysis to the full database of tones, future work seeks to independently characterize the instantaneous frequency and amplitude trajectories of individual overtones of a note. Finally, these statistics may be used to develop a parametric model of randomness in vibrato that could be used for resynthesis applications.

References

- [1] Papich, G. and Rainbow, E., "A pilot study of performance practices of twentieth-century musicians," *Journal of Research in Music Education*, 22(1), pp. 24–34, 1974, doi:<https://doi.org/10.2307/3344615>.
- [2] Gough, C. E., "Measurement, modelling and synthesis of violin vibrato sounds," *Acta Acustica united with Acustica*, 91(2), pp. 229–240, 2005.

-
- [3] Fletcher, N. H., “Acoustical correlates of flute performance technique,” *The Journal of the Acoustical Society of America*, 57(1), pp. 233–237, 1975.
- [4] Fletcher, H. and Sanders, L. C., “Quality of violin vibrato tones,” *The Journal of the Acoustical Society of America*, 41(6), pp. 1534–1544, 1967, doi:<https://doi.org/10.1121/1.1910516>.
- [5] De Cheveigné, A. and Kawahara, H., “YIN, a fundamental frequency estimator for speech and music,” *The Journal of the Acoustical Society of America*, 111(4), pp. 1917–1930, 2002, doi:<https://doi.org/10.1121/1.1458024>.
- [6] Peeters, G., “A large set of audio features for sound description (similarity and classification) in the CUIDADO project,” 2004.
- [7] Smith III, J. O., *Spectral audio signal processing*, W3K publishing, 2011.
- [8] Gardner, W. A., “A unifying view of coherence in signal processing,” *Signal Processing*, 29(2), pp. 113–140, 1992.
- [9] Fritts, L., “University of Iowa musical instrument samples,” on-line at <http://theremin.music.uiowa.edu/MIS.html>, 1997.
- [10] Opolko, F. J. and Wapnick, J., *McGill University Master Samples [kit]: MUMS*, McGill University, Faculty of Music, 1989.
- [11] Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R., “RWC music database: Music genre database and musical instrument sound database,” 2003.